The R Statistical Computing Environment
Basics and Beyond
Survival Analysis: Exercises

John Fox
(McMaster University)
ICPSR/Berkeley

2016

1. The file `Henning.txt` (on the workshop web site) contains data from yet another study of criminal recidivism, this one by Henning and Frueh (1996), who followed 194 inmates released from a medium-security prison to a maximum of three years from the day of their release; during the period of the study, 106 of the released prisoners were rearrested (depressing, eh?). The data set, which is discussed by Judith Singer and John Willet in *Applied Longitudinal Data Analysis* (Oxford, 2003), contains the following variables (using the names employed by Singer and Willet):

   - `months`: The time of re-arrest in months (but measured to the nearest day).
   - `censor`: A dummy variable coded `1` for censored observations and `0` for uncensored observations. Note that this is the opposite of our usual convention, so in R survival time and censoring should be specified as `Surv(months, censor == 0)`; be careful to use the double-equals sign for testing equality!
   - `personal`: A dummy variable coded `1` for prisoners with a record of crime against persons and `0` otherwise.
   - `property`: A dummy variable coded `1` for prisoners with a record of crime against property and `0` otherwise.
   - `cage`: "Centered" age in years at time of release — that is, age − average age.

   (a) Compute and graph the Kaplan-Meier estimate of the survival function for all of the data.

   (b) Compute and graph separate survival curves for those with and without a record of crime against persons; test for differences between the two survival functions.

   (c) Compute and graph separate survival curves for those with and without a record of crime against property; test for differences between the two survival functions.

2. Continuing with the Henning and Frueh data, fit a Cox regression of time to re-arrest on the covariates `personal`, `property`, and `cage`.

   (a) Determine by a Wald test whether each estimated coefficient is statistically significant.

   (b) Interpret each of the estimated Cox-regression coefficients.